SOLVING BLOCK BANDED BLOCK TOEPLITZ SYSTEMS WITH STRUCTURED BLOCKS: ALGORITHMS AND APPLICATIONS

DARIO ANDREA BINI AND BEATRICE MEINI*

Abstract. After motivating the interest and the role of Toeplitz matrices by means of several examples of applications, we focus our interest on the solution of block banded block Toeplitz systems endowed with a two-level structure. In particular we consider block tridiagonal block Toeplitz systems where the blocks are Toeplitz themselves or have a Kronecker product structure, or are Frobenius-like matrices. We introduce some computational tools like the orthogonal displacement representation and the concept of approximate displacement rank and combine them with other well known tools for the design of effective algorithms for the solution of structured problems. Applications to polynomial factorization, solution of queueing models, and image restoration are shown. Finally the concept of approximate displacement rank is used for the design of effective algorithms for Toeplitz-like matrix inversion based on Newton's iteration.

1. Matrix structures and their role. The mathematical modeling of problems of the real world often leads to solving linear systems of equations endowed with specific structures and with very large dimensions. Generally the matrices involved in these models are defined by few parameters that characterize all their entries through a given pattern.

Certain structures are very frequent and reflect specific features that are common to problems arising in diverse fields of applied mathematics and engineering. Toeplitz matrices are an example of this fact. A matrix is Toeplitz if it has constant entries along each diagonal. This kind of matrices, with their block versions, arises whenever properties of *shift invariance* are satisfied by some function in the model. They are encountered, in particular, in fields like image processing, queueing theory, computer algebra and in the numerical solution of differential equations where the shift invariance takes different forms.

Indeed, the stronger is the structure, the more likely efficient solution algorithms can be designed. A challenging problem is to analyze and exploit, as most as possible, the specific structures that characterize a problem in order to devise tools for the design of efficient solution algorithms.

Very often such structures are accompanied by further properties, like banded patterns or Kronecker product patterns that make the problem still more interesting. For block matrices we may have structures at two (or more) different levels: an inner structure, typical of each block, and an outer structure proper of the block matrix. Exploiting both structures is a nontrivial task as it could seem at first.

In fact, it may happen that algorithms designed for the exploitation of the outer structure destroy the inner one and vice-versa.

Concerning Toeplitz computation there is a wide and consolidated literature that provides a lot of tools for the design of efficient algorithms.

In this paper we first describe some examples of applications of (block) Toeplitz matrices and relate the specific structures encountered in this analysis with the peculiar features of the problem. Then we present some recent algorithms, based on the Toeplitz matrix technology, for the solution of certain structured linear systems. In particular, we consider matrices with a two-level structure, like block banded block Toeplitz matrices with banded Toeplitz blocks, or with blocks having a Kronecker

* Dipartimento di Matematica, Università di Pisa, via Buonarroti 2, I-56127 Pisa, Italy. E-mail: bini@dm.unipi.it, meini@dm.unipi.it

product pattern. These problems arise in queueing theory, in image processing and in polynomial computations.

We provide some new techniques for dealing with such problems and discuss some related issues. In particular we introduce the concepts of orthogonal displacement representation, useful to improve the numerical stability of the computation, and the concept of approximate displacement rank. The latter tool is particularly suitable for matrices having a small Euclidean distance from the set of matrices with low displacement rank.

We show applications of these concepts to improving and extending known techniques, like cyclic reduction, to multilevel Toeplitz matrices. We also apply approximate displacement rank for the acceleration of Newton's iteration for inverting Toeplitz-like matrices.

The paper is organized in the following way. This section continues with some examples where the Toeplitz structure is fundamental in the modeling of the mathematical problem. Section 2 is devoted to the description of some tools for structure analysis, in particular we introduce orthogonal displacement representations, approximate displacement rank and recall the main features of cyclic reduction.

In section 3 we present the solution of diverse problems; specifically, the solution of banded Toeplitz systems with application to polynomial factorization; the solution of two queueing models; the solution of block banded block Toeplitz systems encountered in image processing; the inversion of Toeplitz-like matrices by means of Newton's iteration and approximate displacement rank.

1.1. Structures in queueing problems. In the numerical solution of Markov chains, we need to solve (infinite) linear systems $Q\boldsymbol{x} = \boldsymbol{b}$, where $Q = I - P$ and P is a stochastic, or substochastic matrix. The matrix P, whose (i, j) entry $p_{i,j}$ is the probability of transition from the state i to the state j of the system in one unit of time, can be strongly structured, according to the specific model.

In many problems of the real world (eg. network communication, queueing problems), the transition probabilities $p_{i,j}$ are shift invariant, i.e., $p_{i+h,j+h} = p_{i,j}$. This is a very natural situation in queueing models where the probability that the number j of customers waiting in the system is changed into i, in a time unit, depends only on the difference $i - j$, that coincides with the difference between the number of new arrivals and the number of customers served in the unit time. Evidently this shift invariance property leads to Toeplitz structured matrices.

In many models it is not allowed that in a time unit the number of waiting customers is increased by an additive number larger than k or smaller than $-h$, for positive h, k. This property, i.e., $p_{i,j} = 0$ for $j - i > k$, or $j - i < -h$, leads to a lower, or upper, generalized Hessenberg structure.

For systems with more than one queue, the scalar Toeplitz, lower/upper (generalized) Hessenberg structure of the transition probability matrix is replaced with a block Toeplitz, block lower/upper (generalized) Hessenberg structure.

In particular, in the case of QBD (Quasi-Birth-Death) processes [26], $Q = I - P$ is a block tridiagonal block Toeplitz matrix except for the block in position (1,1), i.e.,

$$Q = \begin{bmatrix} \widehat{A} & C & & \bigcirc \\ B & A & C & \\ & B & A & \ddots \\ \bigcirc & & \ddots & \ddots \end{bmatrix}, \tag{1.1}$$

where A, $\widehat{A}$, B, C are $m \times m$ matrices.

The block entries A, $\widehat{A}$, B, C can have also an inner structure; for example, for PH/PH/1 queues [26, 3], that are a particular case of QBD processes, the blocks, having size $m = m_1 m_2$, are given by

$$A = T \oplus S, \quad C = \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I_{m_2}, \quad B = I_{m_1} \otimes \boldsymbol{S}^0 \boldsymbol{\beta}, \tag{1.2}$$

where $T \in \boldsymbol{R}^{m_1 \times m_1}$, $S \in \boldsymbol{R}^{m_2 \times m_2}$, $\boldsymbol{\alpha} \in \boldsymbol{R}^{1 \times m_1}$, $\boldsymbol{\beta} \in \boldsymbol{R}^{1 \times m_2}$, $\boldsymbol{T}^0$ and $\boldsymbol{S}^0$ are such that $T\boldsymbol{e} + \boldsymbol{T}^0 = 0$ and $S\boldsymbol{e} + \boldsymbol{S}^0 = 0$, $\boldsymbol{e} = (1, 1, \ldots, 1)^T$, the symbols $\otimes$ and $\oplus$ denote the Kronecker product and the Kronecker sum of matrices, respectively, and I_r denotes the $r \times r$ identity matrix.

Another interesting class of queueing problems is the one modeled by Non-skip-free M/G/1-type Markov chains [16, 8]; in this case the matrix Q is a block Toeplitz matrix in generalized block Hessenberg form, i.e.,

$$Q = \begin{bmatrix} A_k & A_{k-1} & \ldots & A_0 & & \bigcirc \\ A_{k+1} & A_k & A_{k-1} & \ldots & A_0 & \\ A_{k+2} & A_{k+1} & A_k & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}, \tag{1.3}$$

where A_i, $i = 0, 1, \ldots$, are $m \times m$ matrices. In the applications, the above matrix is approximated with a banded matrix, i.e., it is assumed that $A_i = 0$, for $i > N$, for a suitable integer N. The matrix Q can be reblocked into $k \times k$ block matrices thus obtaining a block Hessenberg block Toeplitz matrix with structured block entries. Hence, solving $Q\boldsymbol{x} = \boldsymbol{b}$ can be reduced to the problem of the solution of a system with a block Toeplitz block Hessenberg matrix with structured blocks.

Infinite block Toeplitz block Hessenberg systems can be reduced to block tridiagonal block Toeplitz systems where the matrix, of the kind (1.1), has infinite blocks, by means of a procedure devised by V. Ramaswami in [28] that consists in adding new auxiliary variables and new equations. The infinite blocks A, B, C that are obtained in this way have a strong structure that is related to block Frobenius matrices. More precisely, if the original block Hessenberg block Toeplitz matrix is defined by (1.3) with $k = 1$, then we have

$$A = \begin{bmatrix} A_1 & A_2 & A_3 & \ldots \\ & I & & \\ & & I & \\ & & & \ddots \end{bmatrix}, \quad B = \begin{bmatrix} A_0 & 0 & \ldots \\ 0 & 0 & \ldots \\ \vdots & \vdots & \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & 0 & \ldots \\ -I & 0 & \ddots \\ & -I & \ddots \\ & & \ddots \end{bmatrix}.$$

Once again we encounter a matrix with a two-level structure where the external block Toeplitz structure is combined with the internal block Frobenius structure.

1.2. Structures in polynomial computations. Polynomials, power series and (infinite) Toeplitz matrices are closely related. If $a(x) = \sum_{i=0}^{+\infty} a_i x^i$, $b(x) =$

$\sum_{i=0}^{+\infty} b_i x^i$, $c(x) = \sum_{i=0}^{+\infty} c_i x^i$ are formal power series such that $c(x) = a(x)b(x)$, then the relations that express the coefficient of $c(x)$ with those of $a(x)$ and $b(x)$ are synthesized by

$$\begin{bmatrix} c_0 \\ c_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} a_0 & & \\ a_1 & a_0 & \\ \vdots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \end{bmatrix}.$$

The same relation holds for polynomials assuming that the coefficients of x^i in the power series $a(x)$, $b(x)$ and $c(x)$ are zero when i is larger than the degree. The Toeplitz structure of the matrix in the above relation comes from the fact that multiplying a power series by the variable x corresponds to multiplying its vector coefficients by the Toeplitz matrix $Z = (z_{i,j})$, where $z_{i+1,i} = 1$, $z_{i,j} = 0$ elsewhere.

For matrix power series, i.e., power series where the coefficients are $m \times m$ matrices the above relation still holds where the matrix involved turns into a block Toeplitz matrix.

This correlation between Toeplitz matrices and power series allows one to express almost any polynomial or power series computation (like gcd, lcm, Euclidean scheme, power series arithmetic, etc) in terms of Toeplitz computations and vice-versa (see [13], [7], [2]).

Recently, this correlation has been used for devising an efficient algorithm for polynomial and power series factorization [4].

Let $p(z) = \sum_{i=0}^{2m} p_i z^i$ be a polynomial of degree (at most) $2m$, having zeros z_i, $i = 1, \ldots, 2m$, such that $|z_1| \leq \ldots \leq |z_m| < |z_{m+1}| \leq \ldots \leq |z_{2m}|$, and denote $\boldsymbol{f}^T = (f_0, \ldots, f_{m-1})$ the coefficient vector of the polynomial $f(z) = \prod_{i=1}^{m}(z - z_i) = \sum_{i=0}^{m} f_i z^i$, $f_m = 1$.

Consider the $n \times n$ banded Toeplitz matrix $Q_n = (q_{i,j})$, $q_{i,j} = p_{i-j+m}$ if $0 \leq i - j + m \leq 2m$, $q_{i,j} = 0$, otherwise. A consequence of Koenig's theorem [19] implies that for n large enough Q_n is nonsingular, moreover, denoting $\boldsymbol{w}^{(n)}$ the vector made up by the first m components of the first column of Q_n^{-1}, it holds $\boldsymbol{f} = -U\boldsymbol{w}^{(n)} + O(\sigma^{mn})$, where $\sigma = |z_m/z_{m+1}|$ and U is the $m \times m$ upper triangular Toeplitz matrix defined by its first row $p_{2m}, \ldots, p_{m+1}$ (compare with [4]).

Thus the problem of approximating a factor of a polynomial is reduced to compute the first m entries of the first column of the inverse of Q_n for a large enough value of n. This computation can be performed by an algorithm that provides the above values for $n = m2^i$, $i = 1, 2, 3, \ldots$, and where the cost of switching from $n = m2^i$ to $n = m2^{i+1}$ is $O(m \log^2 m)$ arithmetic operations.

1.3. Structures in image processing. In many problems of image restoration one of the main issues is to recover an image from a blurred one, where the blur operator is linear, space (shift) invariant and has a local action. More precisely, if we represent images as $m \times n$ arrays X of reals, each of them representing a pixel (picture element) of the image, the discrete blur operator can be described as $B = L(X)$, where the blurred image $B = (b_{i,j})_{i=1,m,j=1,n}$ is related to the original image $X = (x_{i,j})_{i=1,m,j=1,n}$ by

$$b_{i,j} = \sum_{h=-M}^{M} \sum_{k=-M}^{M} \alpha_{h,k} x_{i+h,j+k}, \tag{1.4}$$

(here we assume that $x_{i,j} = 0$ if i or j is out of range) and the Point Spread Function (PSF) matrix $\boldsymbol{\alpha} = (\alpha_{i,j})_{i,j=-M,\ldots,M}$ is such that $\sum_{i,j} \alpha_{i,j} = 1$, $\alpha_{i,j} \geq 0$.

The entry $\alpha_{i-h,j-k}$ represents the influence in the pixel (i,j) of the blurred image of a unit source placed in the pixel (h,k) of the original image.

Rearranging the entries of X and B column-wise as vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ in (1.4) we obtain the linear system $Q\boldsymbol{x} = \boldsymbol{b}$, where the matrix Q is an $n \times n$ block Toeplitz block banded matrix with $m \times m$ banded Toeplitz blocks. More precisely, Q is a block $(2M+1)$-diagonal matrix with $(2M+1)$-diagonal blocks. Moreover the blocks on the i-th block diagonal, $i = -M, \ldots, M$, coincide with the $(2M+1)$-diagonal Toeplitz matrix having on the j-th diagonal the entry $\alpha_{i,j}$, for $j = -M, \ldots, M$. The block Toeplitz structure is derived by the shift invariance properties of the PSF, i.e., the mask $\boldsymbol{\alpha}$ is independent of the specific pixel to which is applied. The block banded structure is due to the locality of the PSF.

1.4. Structures in differential equations. Differential equations with boundary conditions can be efficiently treated by means of numerical techniques. For instance, the Sturm-Liouville problem

$$\begin{cases} -(a(x)u'(x))' = f(x) \\ u(0) = \alpha_0, \quad u(1) = \alpha_1 \end{cases}$$

can be numerically solved by replacing the unknown function $u(x)$ with the unknown vector $\boldsymbol{u} = (u_i)$, where $u_i = u(x_i)$ and $x_i = i/(n+1)$, $i = 1, \ldots, n$. The differential operator, discretized by means of the finite differences technique, is reduced to a tridiagonal matrix $T = (t_{i,j})$, $t_{i,i} = (a_i + a_{i-1})/(n+1)$, $t_{i,i+1} = t_{i+1,i} = -a_i/(n+1)$ and the vector $\boldsymbol{u}$ solves the system $T\boldsymbol{u} = \boldsymbol{f} + O(h^2)$, $h = 1/(n+1)$, where $\boldsymbol{f} = (f_i)$, $f_i = f(x_i)$, $i = 2, \ldots, n-1$, $f_1 = f(x_1) + (n+1)^2\alpha_0$, $f_n = f(x_n) + (n+1)^2\alpha_1$.

It is immediate to verify that if $a(x)$ is constant, that is if the differential operator reduces to the second derivative, then the matrix T is Toeplitz; for a generic $a(x)$ the matrix T is not Toeplitz but is "locally Toeplitz" in the sense of [29]. For derivatives of higher order with constant coefficients the matrix T is still banded Toeplitz and its band-width is proportional to the order of the derivative. For a linear differential operator of functions in m variables with constant coefficients, like the Laplace operator, the matrix that provides the numerical discretization over $[0,1]^m$ is a multilevel block Toeplitz matrix with a multilevel band.

The band Toeplitz structure of the matrices discretizing linear differential operators with constant coefficients is essentially due to the shift invariance property of derivatives expressed by the following relation

$$D^k(f(x+h)) = g(x+h), \quad g(x) = D^k(f(x)),$$

valid for the differential operator $D(f(x)) = f'(x)$.

2. Some tools for structure analysis. In this section we recall some basic tools for the design of efficient algorithms for solving certain structured systems. Throughout this section we assume that all the matrices have real entries unless explicitly specified.

2.1. Displacement rank. The concept of displacement operators and displacement rank, introduced in [22, 21] and elaborated in [17, 13], is a powerful tool for dealing with Toeplitz matrices. Here we recall the main ideas and results.

Let us consider the operator

$$\Delta(A) = AZ - ZA, \quad Z = (z_{i,j}), \quad z_{i+1,i} = 1, \quad z_{i,j} = 0, \text{ otherwise}, \tag{2.1}$$

applied to $n \times n$ matrices A, and define the *displacement rank* of A, associated with Δ, as $\mathrm{Drk}(A) = \mathrm{rank}(\Delta(A))$. A Toeplitz matrix has displacement rank at most 2, moreover, $\mathrm{Drk}(A)$ is a sort of measure of the distance of A from the algebra of lower triangular Toeplitz matrices, for which the displacement rank is zero.

Define $L(\boldsymbol{a})$ the $n \times n$ lower triangular Toeplitz matrix having $\boldsymbol{a}$ as first column. It holds [13]

$$\Delta(A) = \sum_{i=1}^{r} \boldsymbol{u}_i \boldsymbol{v}_i^T \quad \Leftrightarrow A = L(A\boldsymbol{e}_1) + \sum_{i=1}^{r} L(\boldsymbol{u}_i) L^T(Z\boldsymbol{v}_i), \tag{2.2}$$

where $\boldsymbol{u}_i$, $\boldsymbol{v}_i$, $i = 1, \ldots, r$, are n-dimensional vectors, $\boldsymbol{e}_1$ denotes the first column of the identity matrix. The relation in the right-hand side of (2.2) is called the *displacement representation of* A and the vectors $\boldsymbol{a} = A\boldsymbol{e}_1$, $\boldsymbol{u}_i$, $\boldsymbol{v}_i$, $i = 1, \ldots, r$ are called a *set of generators for the displacement representation of* A. Equivalent representations can be given in terms of different operators, see for instance [20, 13].

A simple consequence of the relation

$$\Delta(A^{-1}) = -A^{-1}\Delta(A)A^{-1}, \tag{2.3}$$

that can be directly derived from (2.1), is that $\mathrm{Drk}(A^{-1}) = \mathrm{Drk}(A)$ and the following displacement representation of A^{-1} holds

$$A^{-1} = L(A^{-1}\boldsymbol{e}_1) - \sum_{i=1}^{r} L(A^{-1}\boldsymbol{u}_i) L^T(ZA^{-T}\boldsymbol{v}_i). \tag{2.4}$$

The representations (2.2), (2.4) allow us to make elementary operations among Toeplitz-like matrices at a low cost by means of FFT. In particular, computing the product of a matrix with displacement rank r and a vector costs $O(rn \log n)$ arithmetic operations (ops), inverting a strongly nonsingular matrix with displacement rank r costs $O(r^2 n \log^2 n)$ ops by using a superfast algorithm like the Bitmead Anderson algorithm [14].

It is a simple matter to show that $\Delta(AB) = A\Delta(B) + \Delta(A)B$, therefore it holds $\mathrm{Drk}(AB) \leq \mathrm{Drk}(A) + \mathrm{Drk}(B)$. Similarly we have $\mathrm{Drk}(A + B) \leq \mathrm{Drk}(A) + \mathrm{Drk}(B)$.

2.2. Orthogonal displacement representation. It is worth to point out that the set of generators for the displacement representation of A is not unique. Indeed, observe that, if $\Delta(A) = UV^T$, for $U, V \in \boldsymbol{R}^{n \times r}$ then $\Delta(A) = U'V'^T$ for $U' = US$, $V' = VS^{-T}$, where $S \in \boldsymbol{R}^{r \times r}$ is any nonsingular matrix. Therefore another displacement representation can be derived by the latter decomposition.

If the product of the norms of the columns of U and the corresponding columns of V is much larger than the norm of A, then computing the matrix vector product $A\boldsymbol{x}$ by means of the displacement representation formula (2.2) may lead to large numerical cancellation errors.

This drawback can be overcome by using the concept of orthogonal displacement representation.

If $\Delta(A) = U\Sigma V^T$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, U, $V \in \boldsymbol{R}^{n \times r}$, is the singular values decomposition of $\Delta(A)$, then the quadruple $(\boldsymbol{a}, U, \boldsymbol{\sigma}, V)$ defines a set of generators, where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_r)$, such that $A = L(\boldsymbol{a}) + \sum_{i=1}^{r} \sigma_i L(\boldsymbol{u}_i) L(Z\boldsymbol{v}_i)^T$, where $\boldsymbol{u}_i = U\boldsymbol{e}_i$, $\boldsymbol{v}_i = V\boldsymbol{e}_i$, $i = 1, \ldots, r$, and $\boldsymbol{e}_i$ denotes the i-th column of the $n \times n$ identity matrix. We refer to this particular set of generators as to the *orthogonal generators for the displacement representation of* A.

Due to the uniqueness of the SVD of a matrix the set of orthogonal generators of a matrix is unique.

Observe that for the Euclidean norm $||\cdot||$ we have $||\Delta(A)|| = ||AZ - ZA|| \leq 2||A||$, and for the orthogonal set of generators it holds $||U|| = ||V|| = 1$, $\sigma_1 = ||\Delta(A)||$, whence $\sigma_1 \leq 2||A||$. This fact overcomes the possibility of numerical cancellation.

If A is such that $A^T = JAJ$, like in the case of Toeplitz matrices, where J is the permutation matrix having 1 in the anti-diagonal, there exists an additional structure of $\Delta(A)$. In fact it can be easily proved that $\Delta(A)J$ is skew-symmetric.

For skew-symmetric matrices we have the following technical result:

LEMMA 2.1. *Let $A \in \boldsymbol{R}^{n\times n}$ be such that $A = -A^T$ and $\mathrm{rank}(A) = r$. Then $r = 2h$ is even and there exist $W \in \boldsymbol{R}^{n\times r}$, $D \in \boldsymbol{R}^{h\times h}$ such that $W^TW = I_r$, $D = \mathrm{diag}(d_1, \ldots, d_h)$ and*

$$A = W(D \otimes S)W^T, \quad S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

moreover $\pm \boldsymbol{i} d_j$ are the nonnull eigenvalues of A, where $\boldsymbol{i}$ is the complex unit such that $\boldsymbol{i}^2 = -1$, and $|d_j| = \sigma_{2j-1} = \sigma_{2j}$, $j = 1, \ldots, h$, where $\sigma_1, \ldots, \sigma_r$ are the singular values of A.

Proof. From the Schur normal form deduce that a skew symmetric matrix A is such that $A = Q\widehat{D}Q^H$, for $Q^HQ = I$, $\widehat{D} = \mathrm{diag}(d_j)$ where $d_{2j-1} = \theta_j \boldsymbol{i}$, $d_{2j} = -\theta_j \boldsymbol{i}$, $\theta_j \in \boldsymbol{R}$, $|\theta_j| = \sigma_{2j-1} = \sigma_{2j}$, $j = 1, \ldots, h$, $d_j = 0$, $j = 2h+1, \ldots, n$, and A^H denotes the transpose conjugate of A. Therefore, only an even number of eigenvalues can be nonzero. Denote $\boldsymbol{q}_j = Q\boldsymbol{e}_j$; since A is real then $\boldsymbol{q}_{2j} = \bar{\boldsymbol{q}}_{2j-1}$; apply the equation

$$\begin{bmatrix} \theta\boldsymbol{i} & 0 \\ 0 & -\theta\boldsymbol{i} \end{bmatrix} = \frac{\theta}{2} \begin{bmatrix} 1 & -\boldsymbol{i} \\ 1 & \boldsymbol{i} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \boldsymbol{i} & -\boldsymbol{i} \end{bmatrix}$$

and conclude that the claim is true where the columns of W are given by $(\boldsymbol{q}_{2j-1} + \bar{\boldsymbol{q}}_{2j-1})\sqrt{2}$ and $(\boldsymbol{q}_{2j-1} - \bar{\boldsymbol{q}}_{2j-1})\boldsymbol{i}\sqrt{2}$, that are orthonormal. □

As a consequence, we have the following fundamental property:

THEOREM 2.2. [Displacement representation of a persymmetric matrix]. *If the persymmetric matrix A has displacement rank r with respect to Δ, then $r = 2h$ and there exist matrices $W \in \boldsymbol{R}^{n\times r}$, $D = \mathrm{diag}(d_1, \ldots, d_h)$ such that $W^TW = I_r$, $|d_1| \geq |d_2| \geq \ldots \geq |d_h|$, $|d_j| = \sigma_{2j-1} = \sigma_{2j}$, $j = 1, \ldots, h$, where $\sigma_1, \ldots, \sigma_r$ are the singular values of $\Delta(A)$, and*

$$\Delta(A) = W(D \otimes S)W^TJ, \quad S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Therefore,

$$A = L(\boldsymbol{a}) + \sum_{i=1}^{h} d_i (L(\boldsymbol{w}_{2i-1})L^T(ZJ\boldsymbol{w}_{2i}) - L(\boldsymbol{w}_{2i})L^T(ZJ\boldsymbol{w}_{2i-1})) \tag{2.5}$$

where $\boldsymbol{a} = A\boldsymbol{e}_1$, $\boldsymbol{w}_i = W\boldsymbol{e}_i$, $i = 1, \ldots, r$. The orthogonal generator for A is completely defined by the triple $(\boldsymbol{a}, W, \boldsymbol{d})$ where $\boldsymbol{d} = (d_i)_{i=1,\ldots,h}$.

This fact can be exploited for reducing the computational cost of multiplying the matrix A and a vector. Indeed, it is easy to prove that, given the FFT associated with $\boldsymbol{x}$ and $\boldsymbol{y}$, for the computation of the products $L(\boldsymbol{w})\boldsymbol{x}$ and $L(ZJ\boldsymbol{w})\boldsymbol{y}$ only the FFT associated with $\boldsymbol{w}$ is sufficient.

2.3. Approximate displacement rank. In certain situations it may be useful to measure the distance of a matrix A from the set of triangular Toeplitz matrices in an approximate way. In fact, it may happen that a matrix A is "very far" from the set of triangular Toeplitz matrices since its displacement rank is large (say, n), but its Euclidean distance is small, say, of the order of the machine precision. This occurs, for instance, for the matrix

$$A = T + E$$

where T is triangular Toeplitz and for the Euclidean norm $||E||$ it holds $||E|| < \epsilon$. In this case representing A in terms of its "full" displacement representation is not suitable and is more convenient to replace A with T which is a very accurate approximation (if ϵ is small enough) and can be efficiently represented in terms of its displacement representation.

DEFINITION 2.3. *For a given $\epsilon > 0$ define the ϵ-displacement rank of a matrix A as* $\mathrm{Drk}_\epsilon(A) = \min_{||E|| \le \epsilon} \mathrm{rank}(\Delta(A) + E)$, *and the relative ϵ-displacement rank as* $\mathrm{Rdrk}_\epsilon(A) = \min_{||E|| \le \epsilon ||A||} \mathrm{rank}(\Delta(A) + E)$.

If $r = \mathrm{Drk}_\epsilon(A)$ ($r = \mathrm{Rdrk}_\epsilon(A)$), and if $\Delta(A) = \sum_{i=1}^{r} \boldsymbol{u}_i \boldsymbol{v}_i^T + E$, with $||E|| \le \epsilon$ ($||E|| \le \epsilon ||A||$), we say that the matrix

$$L(A\boldsymbol{e}_1) + \sum_{i=1}^{r} L(\boldsymbol{u}_i) L^T(Z\boldsymbol{v}_i) \tag{2.6}$$

is an *ϵ-displacement (relative ϵ-displacement) representation* of A.

The following properties hold.

THEOREM 2.4. *Let $\Delta(A) = U\Sigma V^T$ be the singular value decomposition of $\Delta(A)$, where $\sigma_1 \ge \ldots \ge \sigma_k > 0$ are the singular values of $\Delta(A)$. Let $\sigma_1 > \epsilon \ge \sigma_k$; then $\mathrm{Drk}_\epsilon(A) = r$ if and only if $\sigma_r > \epsilon \ge \sigma_{r+1}$. Similarly, let $\sigma_1 > \epsilon ||A|| \ge \sigma_k$; then $\mathrm{Rdrk}_\epsilon(A) = r$ if and only if $\sigma_r > \epsilon ||A|| \ge \sigma_{r+1}$.*

Proof. We prove that $\sigma_{r+1} \le \epsilon$ if and only if $\mathrm{Drk}_\epsilon(A) \le r$. This is sufficient to prove the theorem. Let $\boldsymbol{u}_i$, $\boldsymbol{v}_i$, $i = 1, \ldots, k$, be the i-th column of U and V, respectively, that is

$$\Delta(A) = \sum_{i=1}^{r} \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^T + \sum_{i=r+1}^{k} \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^T. \tag{2.7}$$

Define $E = \sum_{i=r+1}^{k} \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^T$, then $||E|| = \sigma_{r+1}$. Thus, if $\sigma_{r+1} \le \epsilon$, we have $\mathrm{Drk}_\epsilon(A) \le \mathrm{Drk}(\Delta(A) - E) = r$. Vice-versa, if $\mathrm{Drk}_\epsilon(A) = r$ then there exists a matrix E, with $||E|| \le \epsilon$, such that $\mathrm{rank}(\Delta(A) + E) = r$. Define $\theta_r = \min_{\{X:\ \mathrm{rank}(X) = r\}} ||\Delta(A) - X||$; since $\theta_r = \sigma_{r+1}$ and $\sigma_{r+1} = ||\Delta(A) - \sum_{i=1}^{r} \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^T||$, it follows that $\sigma_{r+1} = \theta_r \le ||E||$. □

THEOREM 2.5. *Let $\sigma_1, \ldots, \sigma_k$ be the singular values of $\Delta(A)$, $\sigma_1 > \epsilon \ge \sigma_k$, $r = \mathrm{Drk}_\epsilon(A) \le \mathrm{Drk}(A) = k$, and let A_ϵ be an ϵ-displacement representation of A. Then it holds*

$$||A - A_\epsilon|| \le n \sum_{i=r+1}^{k} \sigma_i \le n(k-r)\epsilon.$$

Similarly, if $r = \mathrm{Rdrk}_\epsilon(A)$ and if A_ϵ is a relative ϵ-displacement representation of A, then it holds

$$\frac{||A_\epsilon - A||}{||A||} \le n \sum_{i=r+1}^{k} \sigma_i \le n(k-r)\epsilon.$$

Proof. Let $\Delta(A) = U\Sigma V^T$ be the singular values decomposition of $\Delta(A)$. From Theorem 2.4 and (2.7) it holds $A - A_\epsilon = \sum_{i=r+1}^{k} \sigma_i L(\boldsymbol{u}_i) L(Z\boldsymbol{v}_i)^T$, where $||\boldsymbol{u}_i|| = ||\boldsymbol{v}_i|| = 1$. Since for any vector $\boldsymbol{u}$ it holds $||L(\boldsymbol{u})|| = ||L(\boldsymbol{u})^T||$ and $||L(\boldsymbol{u})|| \le \sum_{i=1}^{n} |u_i| \cdot ||Z^i|| = \sum_{i=1}^{n} |u_i| \le \sqrt{n}||u||$, we have $||A - A_\epsilon|| \le n \sum_{i=r+1}^{k} \sigma_i$. The proof is completed since $|\sigma_i| \le \epsilon$ for $r = \mathrm{Drk}_\epsilon(A)$ and $|\sigma_i| \le \epsilon||A||$ for $r = \mathrm{Rdrk}_\epsilon(A)$, $i = r+1, \ldots, k$. □

Conversely, we have the following

THEOREM 2.6. *If A has displacement rank k and B is an approximation of A such that $||A - B|| \le \epsilon$, then $\mathrm{Drk}_{2\epsilon}(B) \le k$. Similarly, if $||A - B|| \le \epsilon||A||$ then $\mathrm{Rdrk}_{2\epsilon}(B) \le k$.*

Proof. It holds $\Delta(B) = \Delta(A) + \Delta(B - A)$, moreover, $||\Delta(B - A)|| \le ||Z(B - A)|| + ||(B - A)Z|| \le 2||B - A|| \le 2\epsilon$ since $||Z|| = 1$. □

THEOREM 2.7. *The following relations extend to the approximate displacement rank the analogous inequalities valid for displacement rank.*

$$
\begin{aligned}
&\mathrm{Drk}_\delta(A+B) \le \mathrm{Drk}_\epsilon(A) + \mathrm{Drk}_\epsilon(B), && \delta = 2\epsilon \\
&\mathrm{Drk}_\delta(AB) \le \mathrm{Drk}_\epsilon(A) + \mathrm{Drk}_\epsilon(B), && \delta = \epsilon(||A|| + ||B||) \\
&\mathrm{Drk}_\delta(A^{-1}) \le \mathrm{Drk}_\epsilon(A), && \delta = ||A^{-1}||^2\epsilon \\
&\mathrm{Rdrk}_\delta(A+B) \le \mathrm{Rdrk}_\epsilon(A) + \mathrm{Rdrk}_\epsilon(B), && \delta = \epsilon\frac{||A||+||B||}{||A+B||} \\
&\mathrm{Rdrk}_\delta(AB) \le \mathrm{Rdrk}_\epsilon(A) + \mathrm{Rdrk}_\epsilon(B), && \delta = 2\epsilon\frac{||A||\cdot||B||}{||AB||} \\
&\mathrm{Rdrk}_\delta(A^{-1}) \le \mathrm{Rdrk}_\epsilon(A), && \delta = \epsilon\mathrm{Cond}(A),
\end{aligned}
\tag{2.8}
$$

where $\mathrm{Cond}(A) = ||A|| \cdot ||A^{-1}||$ *is the spectral condition number of* A.

Proof. Let us prove the second inequality. Assuming $\Delta(A) = S_A + E_A$, $\Delta(B) = S_B + E_B$, where $||E_A|| \le \epsilon$, $||E_B|| \le \epsilon$, from the relation $\Delta(AB) = A\Delta(B) + \Delta(A)B$, we have

$$\Delta(AB) = AS_B + S_AB + E_{AB}, \quad E_{AB} = AE_B + E_AB.$$

Thus $||E_{AB}|| \le \epsilon(||A|| + ||B||)$ and $\mathrm{rank}(AS_B + S_AB) \le \mathrm{rank}(S_A) + \mathrm{rank}(S_B)$. Similarly we do for the remaining inequalities. □

2.4. Cyclic reduction for block tridiagonal block Toeplitz systems. Cyclic reduction is a well known technique originally introduced for the solution of certain systems arising from the discretization of elliptic problems [18], and recently adjusted for the solution of Markov chains [24, 5, 6, 7]. Numerical properties of this method are described in [31, 1]. Here we recall the main properties.

Consider the system

$$Q\boldsymbol{x} = \boldsymbol{b}, \tag{2.9}$$

where Q is a finite block tridiagonal block Toeplitz matrix, of block size 2^q, say, obtained by truncating (1.1), where $\widehat{A} = A$. Partition the vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ into blocks $\boldsymbol{x}_i$, $\boldsymbol{b}_i$ of dimension m, respectively, that is, $\boldsymbol{x} = (\boldsymbol{x}_i)_{i=1,\ldots,2^q}$, $\boldsymbol{b} = (\boldsymbol{b}_i)_{i=1,\ldots,2^q}$. By performing an even-odd permutation of the block rows and block columns in (2.9) we find that

$$\begin{bmatrix} V_0 & U_0 \\ L_0 & W_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_+^{(0)} \\ \boldsymbol{x}_-^{(0)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_+^{(0)} \\ \boldsymbol{b}_-^{(0)} \end{bmatrix} \tag{2.10}$$

where $\boldsymbol{x}_+^{(0)} = (\boldsymbol{x}_{2k})$, $\boldsymbol{x}_-^{(0)} = (\boldsymbol{x}_{2k-1})$, $\boldsymbol{b}_+^{(0)} = (\boldsymbol{b}_{2k})$, $\boldsymbol{b}_-^{(0)} = (\boldsymbol{b}_{2k-1})$,

$$V_0 = W_0 = \begin{bmatrix} A & & \bigcirc \\ & \ddots & \\ \bigcirc & & A \end{bmatrix}, \quad L_0 = \begin{bmatrix} C & & & \bigcirc \\ B & C & & \\ & \ddots & \ddots & \\ \bigcirc & & B & C \end{bmatrix},$$

$$U_0 = \begin{bmatrix} B & C & & \bigcirc \\ & B & \ddots & \\ & & \ddots & C \\ \bigcirc & & & B \end{bmatrix}.$$

Applying one step of block-Gaussian elimination to the 2×2 block system (2.10) yields the equivalent system

$$\begin{cases} (W_0 - L_0 V_0^{-1} U_0)\boldsymbol{x}_-^{(0)} = \boldsymbol{b}_-^{(0)} - L_0 V_0^{-1} \boldsymbol{b}_+^{(0)} \\ \boldsymbol{x}_+^{(0)} = V_0{}^{-1}(\boldsymbol{b}_+^{(0)} - U_0 \boldsymbol{x}_-^{(0)}). \end{cases} \tag{2.11}$$

Denoting

$$Q^{(1)} = W_0 - L_0 V_0^{-1} U_0 \tag{2.12}$$

the Schur complement of V_0, $\boldsymbol{x}^{(1)} = \boldsymbol{x}_-^{(0)}$, $\boldsymbol{b}^{(1)} = \boldsymbol{b}_-^{(0)} - L_0 V_0^{-1} \boldsymbol{b}_+^{(0)}$, the above system is ultimately reduced to solving

$$Q^{(1)} \boldsymbol{x}^{(1)} = \boldsymbol{b}^{(1)}. \tag{2.13}$$

Once $\boldsymbol{x}^{(1)}$ has been computed, the solution of the original system (2.9) can be recovered by means of back substitution through (2.11).

It is easy to verify that $Q^{(1)}$ is a block tridiagonal matrix which, except for the north western corner block, has the block Toeplitz structure. If we apply to system (2.13) the same block even-odd permutation, followed by one step of Gaussian elimination (i.e., one step of cyclic reduction), we obtain a new system, whose matrix $Q^{(2)}$ is a block tridiagonal matrix which, except for the north western corner block, has the block Toeplitz structure. In this way we obtain a sequence of systems $\{Q^{(j)}\boldsymbol{x}^{(j)} = \boldsymbol{b}^{(j)}\}$, of block size 2^{q-j}, $j = 0, \ldots, q$, where $Q^{(0)} = Q$, and

$$Q^{(j)} = \begin{bmatrix} \widehat{A}_j & C_j & & \bigcirc \\ B_j & A_j & C_j & \\ & B_j & A_j & \ddots \\ \bigcirc & & \ddots & \ddots \end{bmatrix}.$$

The block entries A_{j+1}, B_{j+1}, C_{j+1}, $\widehat{A}_{j+1}$, obtained at step $j+1$, that define the matrix $Q^{(j+1)}$, are related to the block entries A_j, B_j, C_j, $\widehat{A}_j$, obtained at step j, by means of the following simple relations:

$$\begin{cases} B_{j+1} = -B_j A_j^{-1} B_j \\ A_{j+1} = A_j - B_j A_j^{-1} C_j - C_j A_j^{-1} B_j \\ C_{j+1} = -C_j A_j^{-1} C_j \\ \widehat{A}_{j+1} = \widehat{A}_j - C_j A_j^{-1} B_j, \end{cases} \tag{2.14}$$

for $A_0 = \widehat{A}_0 = A$, $B_0 = B$, $C_0 = C$. It is useful to express these relations in functional form; for this purpose let us associate with $Q^{(j)}$ the following matrix polynomial

$$\phi_j(z) = C_j + zA_j + z^2 B_j, \quad j = 0, 1, 2, \ldots. \tag{2.15}$$

From (2.12) we may easily deduce that

$$\begin{cases} \phi_{j+1}(z) = zA_j - (C_j + zB_j)A_j^{-1}(C_j + zB_j) \\ \widehat{A}_{j+1} = \widehat{A}_j - C_j A_j^{-1} B_j \end{cases}$$

and that

$$\phi_{j+1}(z^2) = \left(\frac{\phi_j(z)^{-1} - \phi_j(-z)^{-1}}{2z} \right)^{-1} = -\phi_j(z) A_j^{-1} \phi_j(-z).$$

Thus, if we introduce the formal matrix power series $\psi_j(z) = \phi_j(z)^{-1}$, we have

$$\psi_{j+1}(z^2) = \frac{\psi_j(z) - \psi_j(-z)}{2z}, \tag{2.16}$$

that is, the matrix coefficients of $\psi_{j+1}(z)$ are the odd matrix coefficients of $\psi_j(z)$. In other words, the matrix power series $\psi_j(z)$ are determined by matrix coefficients that are suitable coefficients of the matrix power series $\psi(z) = \psi_0(z)$.

Relation (2.16) is useful to prove structural properties of the block entries obtained at each step of cyclic reduction. Indeed, from (2.3) it holds that for any z, $\psi_j(z)$ and $\phi_j(z)$ have the same displacement rank. Unfortunately this does not imply that $\mathrm{Drk}(\psi_j(z)) = \mathrm{Drk}(\psi(z))$ and thus that $\mathrm{Drk}(\phi_j(z)) = \mathrm{Drk}(\psi(z))$ for any j. However, under suitable conditions on $\phi(z)$ the former property is satisfied. This occurs for instance if $\phi(z)$ belongs to the same matrix algebra for any z, or even under weaker conditions. Examples related to banded Toeplitz matrices are discussed in the next section.

It is important to point out that if the matrix polynomials $\phi_j(z)$ has a small displacement rank for any j and z, then the process of cyclic reduction can be implemented at a very low cost with the help of FFT, and with a low memory storage.

Cyclic reduction has very nice convergence properties that allow one to apply this technique even to (bi-)infinite systems. In fact, it can be proved that under suitable conditions $\lim_i B_i = \lim_i C_i = 0$. More precisely we have the following result [11].

THEOREM 2.8. *Assume that A_j is nonsingular for $j \geq 0$. Let ξ_i, $i = 1, \ldots, 2m$, be the zeros of $\phi(z) = C + zA + z^2 B$ and suppose that $|\xi_1| \leq \ldots \leq |\xi_m| < 1 < |\xi_{m+1}| \leq \ldots \leq |\xi_{2m}|$, where we assume that $\phi(z)$ has zeros equal to infinity, if its degree is less than $2m$. Denote $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{2m}$ a set of Jordan chains associated with $\xi_1, \ldots, \xi_{2m}$, respectively, for the matrix polynomial $\phi(z)$. If the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ are linearly independent and if the vectors $\boldsymbol{u}_{m+1}, \ldots, \boldsymbol{u}_{2m}$ are linearly independent then the sequences $||A_j||$, $||{A_j}^{-1}||$ are bounded and for any $\epsilon > 0$ there exists an operator norm $||\cdot||$ such that $||C_j|| = O((|\xi_m| + \epsilon)^{2^j})$, $||B_j|| = O((1/|\xi_{m+1}| + \epsilon)^{2^j})$.*

Proof. (Outline.) For the Jordan chains properties [23, 15] there exist $m \times m$ matrices H and K that solve the matrix equations $C + HA + H^2 B = 0$ and $K^2 C + KA + B = 0$ whose eigenvalues, counting multiplicities, are $\xi_1, \ldots, \xi_m$ and $1/\xi_{m+1}, \ldots, 1/\xi_{2m}$, respectively. It can be proved that [6] at each step j of cyclic reduction, it holds ${C_j A_j}^{-1} + H^{2^j} + H^{2 \cdot 2^j} {B_j A_j}^{-1} = 0$, $K^{2 \cdot 2^j} {C_j A_j}^{-1} + K^{2^j} + {B_j A_j}^{-1} = 0$. Recovering $B_j A_j^{-1}$ from the latter equation and replacing it into the former one,

yields $C_j A_j^{-1} = H^{2^j} R_j$, where R_j is a matrix with bounded norm. Since $\rho(H) = |\xi_m| < 1$, we deduce that for any ϵ there exists an operator norm $||\cdot||$ such that $||H|| \leq |\xi_m| + \epsilon$, that is, $||C_j {A_j}^{-1}|| = O((|\xi_m| + \epsilon)^{2^j})$. Similarly we may do with $B_j {A_j}^{-1}$. Hence, the sequences $||A_j||$, $||{A_j}^{-1}||$ are bounded [31] and therefore $||C_j|| = O((|\xi_m| + \epsilon)^{2^j})$, $||B_j|| = O((1/|\xi_{m+1}| + \epsilon)^{2^j})$. □

If the blocks A, B and C satisfy additional properties we can prove stronger convergence results.

THEOREM 2.9. *Assume that $A = A^T$, A is positive definite and one of the following two conditions holds:*

1. *$B = C^T$, $\gamma(z) = \bar{z}C + A + zB$ is positive definite for $|z| = 1$, $z \in \boldsymbol{C}$;*
2. *$B = -C^T$.*

Then:

- *for $j \geq 1$ $B_j = C_j^T$ and A_j, $\widehat{A}_j$ are positive definite;*
- *$||A_j||$, $||A_j^{-1}||$, $||\widehat{A}_j||$ are bounded from above by a constant;*
- *for any $\epsilon > 0$ there exists a norm $||\cdot||$ such that $||C_j|| = O((\sigma + \epsilon)^{2^j})$, where $\sigma = \max\{|\xi| : \det \gamma(\xi) = 0, |\xi| < 1\}$.*

Moreover, in the hypothesis 1 also $||\widehat{A}_j^{-1}||$ is bounded.

Proof. Assume that condition 1 holds. Consider the block tridiagonal matrix T_n of block size n obtained by truncating (1.1) at block size n with $\widehat{A} = A$. Since the eigenvalues of T_n belong to the interval $[\lambda_1, \lambda_2]$, where $0 < \lambda_1 = \min_{|z|=1} \lambda(\gamma(z))$, $\lambda_2 = \max_{|z|=1} \lambda(\gamma(z))$ [30], and $\lambda(.)$ denotes the set of eigenvalues, then T_n, as well as any Schur complement in T_n, is positive definite. Therefore, A_j, $\widehat{A}_j$, being suitable Schur complements in T_n for suitable values of n [10, 4], have eigenvalues in $[\lambda_1, \lambda_2]$ so that they are positive definite and $||A_j||$, $||A_j^{-1}||$, $||\widehat{A}_j||$, $||\widehat{A}_j^{-1}||$ are bounded. Under hypothesis 1, there exist solutions X, Y to the equations $B^T Y^2 - AY + B = 0$, $BX^2 - AX + B^T = 0$ such that for the spectral radius $\rho(\cdot)$ it holds $\rho(X) = \rho(Y) = \sigma < 1$ [25]. By applying cyclic reduction we obtain that [6, 7] $B_j^T Y^{2 \cdot 2^j} - A_j Y^{2^j} + B_j = 0$, $B_j X^{2 \cdot 2^j} - A_j X^{2^j} + B_j^T = 0$. Recovering B_j^T from the latter and substituting it into the former, we arrive at

$$B_j(I - X^{2 \cdot 2^j} Y^{2 \cdot 2^j}) = A_j(I - X^{2^j} Y^{2^j}) Y^{2^j}.$$

Since $\rho(X) = \rho(Y) < 1$ and $||A_j||$ is bounded, for any $\epsilon > 0$ there exists a norm $||\cdot||$ such that $||B_j|| = O((\sigma + \epsilon)^{2^j})$.

In the assumption 2, since A is nonsingular we may perform one step of cyclic reduction. The matrices A_1, $\widehat{A}_1$, B_1 C_1 that we obtain in this way are such that $B_1 = C_1^T$, A_1, $\widehat{A}_1$ are symmetric and positive definite. Moreover, the function $\gamma_1(z) = zB_1 + A_1 + \bar{z}B_1^T$ is such that $\gamma_1(z) = A + (B - \bar{z}B^T)A^{-1}(B^T - zB)$, that is positive definite for $|z| = 1$. The thesis follows from part 1. □

3. Solving specific structured problems. In this section we analyze the cyclic reduction algorithm, applied to certain block tridiagonal block Toeplitz matrices, having particular block entries.

3.1. The case of banded Toeplitz matrices: application to polynomial factorization. Let $\{a_k\}_{k \in \boldsymbol{Z}}$ be a sequence of real numbers such that $a_k = 0$ for $k < m_2$, and for $k > m_1$, $a_{m_1} a_{m_2} \neq 0$, where $\boldsymbol{Z}$ is the ring of integers, and $m_1, m_2 \in \boldsymbol{Z}$, $m_2 < 0 < m_1$. Let $Q_n = (q_{i,j})$ the matrix of size n having entries $q_{i,j} = a_{i-j}$. Q_n is a banded Toeplitz matrix, with bandwidth $m = \max\{m_1, -m_2\}$. Assuming that

$n = m2^q$, the matrix Q_n can be partitioned into $m \times m$ blocks yielding a block tridiagonal Toeplitz matrix with the structure (1.1), where

$$\widehat{A} = A = \begin{bmatrix} a_0 & a_{-1} & \dots & a_{-m+1} \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{-1} \\ a_{m-1} & \dots & a_1 & a_0 \end{bmatrix},$$

(3.1)

$$C = \begin{bmatrix} a_{-m} & 0 & \dots & 0 \\ a_{-m+1} & a_{-m} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{-1} & \dots & a_{-m+1} & a_{-m} \end{bmatrix}, \quad B = \begin{bmatrix} a_m & a_{m-1} & \dots & a_1 \\ 0 & a_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{m-1} \\ 0 & \dots & 0 & a_m \end{bmatrix}.$$

In order to solve $Q_n \boldsymbol{x} = \boldsymbol{b}$, we apply the cyclic reduction algorithm to the block tridiagonal system, as described in section 2.4. The Toeplitz structure of the blocks A, B, C is generally lost by the blocks A_j, B_j, C_j, $\widehat{A}_j$, obtained in the cyclic reduction process. This fact might lead to think that the cyclic reduction algorithm does not fully exploit the Toeplitz structure of the matrix Q. However it is possible to prove that the displacement rank of the blocks A_j, B_j, C_j, $\widehat{A}_j$ is at most 4. This fact has been proved in [10]. Here we provide a simpler proof of this property that gives more insight to possible extensions and generalizations of it.

Indeed, it is easy to verify that for any z the matrix $\phi(z) = C + zA + z^2B$ is a z-circulant matrix, i.e., belongs to the algebra generated by

$$C_z = \begin{bmatrix} 0 & & & z \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}.$$

Since this set is closed under multiplication and addition, also the inverse $\psi(z) = \phi(z)^{-1}$ is a z-circulant matrix, where $\phi(z)$ is nonsingular. Moreover, since z-circulant matrices are Toeplitz, it follows that $\psi(z)$ is Toeplitz for any z. This implies that the matrices $\psi_j(z)$ are Toeplitz for any z for which they exist, due to (2.16). Therefore $\phi_j(z) = \psi_j(z)^{-1}$, has displacement rank at most 2 for any j and for any z for which $\phi_j(z)$ exists, due to (2.3). In particular we deduce that $C_j = \phi_j(0)$ has displacement rank at most 2, $A_j = (\phi_j(1) + \phi_j(-1))/2$ has displacement rank at most 4, and $B_j = z^2\phi_j(z^{-1})|_{z=0}$ has displacement rank at most 2.

In [10] we provide efficient formulae for the computation of the displacement generators of A_j, B_j, C_j, $\widehat{A}_j$, that lead to $O(m \log^2 m)$ ops the cost for performing one step of cyclic reduction.

For the polynomial factorization problem described in section 1.2, we have to compute the vector $\boldsymbol{w}^{(n)}$, made up by the first m entries of the first column of the inverse of the matrix Q_n, for a large enough value of n, where Q_n is the $n \times n$ banded Toeplitz matrix $Q_n = (q_{i,j})$, such that $q_{i,j} = p_{i-j+m}$ if $0 \le i - j + m \le 2m$, $q_{i,j} = 0$, otherwise. Thus, if $n = m2^q$, we may reblock the matrix Q_n into a block tridiagonal block Toeplitz matrix, with blocks having the structure (3.1), where $a_i = p_{i+m}$, $i = -m, \dots, m$.

In [4] we show that, if we compute the sequence (2.14), $j = 1, \dots, q$, then it holds $\widehat{A}_j^{-1}\boldsymbol{e}_1 = \boldsymbol{w}^{(m2^j)}$. Thus, formulae (2.14) provide the vector $\boldsymbol{w}^{(n)}$, for $n = m2^j$,

$j = 1, 2, 3, \ldots$, and, for the displacement structure of the matrices $\widehat{A}_j$, A_j, B_j, C_j, the cost of switching from $n = m2^j$ to $n = m2^{j+1}$ is $O(m \log^2 m)$ arithmetic operations. Moreover, applying cyclic reduction algorithm leads to a quadratically convergent algorithm, since $\boldsymbol{f} = -U\boldsymbol{w}^{(m2^j)} + O(\sigma^{m2^j})$, $j = 1, 2, \ldots$, where $\sigma = |z_m/z_{m+1}|$.

We may increase the numerical performances of the algorithm by means of the orthogonal displacement representations. In fact it easy to check that A_j, B_j and C_j are persymmetric for any j, thus by using (2.14) and Theorem 2.2 it is easy to compute the orthogonal displacement generators of A_{j+1}, B_{j+1}, C_{j+1} and $\widehat{A}_{j+1}$ given the orthogonal displacement generators of A_j, B_j, C_j, and $\widehat{A}_j$.

For a persymmetric matrix X let us denote $\mathrm{ODG}(X) = (\boldsymbol{x}, W_X, \boldsymbol{d}_X)$, the orthogonal displacement generator of X as given in Theorem 2.2, and let us define $\mathrm{ODG}(X) = (\boldsymbol{x}, U_X, \boldsymbol{\sigma}_X, V_X)$, for a general matrix X, where $U_X \Sigma_X V_X^T$ is the SVD of $\Delta(X)$, $\Sigma_X = \mathrm{diag}(\boldsymbol{\sigma}_X)$ and $\boldsymbol{x}$ is the first column of X. Moreover set $D_X = \mathrm{diag}(\boldsymbol{d}_X) \otimes S$ where $S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

THEOREM 3.1. [Displacement generator in cyclic reduction]. *Let us denote A', B', C' and $\widehat{A}'$ the blocks obtained by applying one step of CR to the blocks A, B, C and $\widehat{A}$, where A, B, C are persymmetric. Then, we have:*

$$\Delta(A') = F_A \begin{bmatrix} D_A & O & O \\ O & O & D' \\ O & D' & O \end{bmatrix} F_A^T J, \quad D' = \begin{bmatrix} -D_B & O & O \\ O & D_A & O \\ O & O & -D_C \end{bmatrix}$$

$$F_A = \left[W_A, W_B, BA^{-1}W_A, BA^{-1}W_C, CA^{-1}W_B, CA^{-1}W_A, W_C\right]$$

$$\Delta(B') = F_B \begin{bmatrix} D_A & O & O \\ O & O & -D_B \\ O & -D_B & O \end{bmatrix} F_B^T J, \quad F_B = \left[BA^{-1}W_A, W_B, BA^{-1}W_B\right]$$

$$\Delta(C') = F_C \begin{bmatrix} D_A & O & O \\ O & O & -D_C \\ O & -D_C & O \end{bmatrix} F_C^T J, \quad F_C = \left[CA^{-1}W_A, W_C, CA^{-1}W_C\right]$$

$$\Delta(\widehat{A}') = F_{\widehat{A}} \mathrm{diag}(\Sigma_{\widehat{A}}, -D_C, D_A, -D_B) G_{\widehat{A}}^T$$

$$F_{\widehat{A}} = \left[U_{\widehat{A}}, W_C, CA^{-1}W_A, CA^{-1}W_B\right], G_{\widehat{A}} = \left[V_{\widehat{A}}, JBA^{-1}W_C, JBA^{-1}W_A, JW_B\right].$$

The above theorem provides a displacement generator of $A, B, C, \widehat{A}$ of the form $U\Sigma V^T$ where the number of columns of U and V is generally much larger than the displacement rank. To compress this redundant displacement representation we may apply the following

ALGORITHM 3.2. *Computing the orthogonal generator for a persymmetric matrix.*

INPUT: Matrices $F \in \boldsymbol{R}^{m \times k}$, $M \in \boldsymbol{R}^{k \times k}$ such that $\Delta(A) = FMF^T J$, $M = -M^T$; $k \geq h$, where $\mathrm{rank}\,\Delta(A) = h$.

OUTPUT: The orthogonal generator of A in the form $(\boldsymbol{a}, W, \boldsymbol{d})$, where $W \in \boldsymbol{R}^{h \times h}$.

COMPUTATION:

STAGE 1: Compute the QR factorization $U = QR$.

STAGE 2: Compute $H = RMR^T$.

STAGE 3: Compute the SVD $H = U\Sigma V^T$, $U, V \in \boldsymbol{R}^{k \times h}$, $\Sigma \in \boldsymbol{R}^{h \times h}$.

STAGE 4: Compute the sign $\gamma_j = \mathrm{sign}(u_{i,2j-1}v_{i,2j})$, $j = 1, \ldots, h$, where i is such that the product is nonzero.

STAGE 4: Set $W = QU$, $d_i = \gamma_i \sigma_{2i-1}$, $i = 1, \ldots, h$, $\boldsymbol{d} = (d_i)$.

For a general matrix A the above algorithm must be modified in the following way

ALGORITHM 3.3. *Computing the orthogonal generator for a general matrix.*

INPUT: Matrices $F_1, F_2 \in \boldsymbol{R}^{m\times k}$, $M \in \boldsymbol{R}^{k\times k}$ such that $\Delta(A) = F_1 M F_2^T$, $k \geq h$, $h = \mathrm{rank}\,\Delta(A)$.

OUTPUT: The orthogonal generator of A in the form $(\boldsymbol{a}, U, \boldsymbol{\sigma}, V)$, $U, V \in \boldsymbol{R}^{m\times h}$.

COMPUTATION:

STAGE 1: Compute the QR factorizations $F_i = Q_i R_i$, $i = 1, 2$.

STAGE 2: Compute $H = R_1 M R_2^T$.

STAGE 3: Compute the SVD $H = G_1 \Sigma G_2^T$, $G_1, G_2 \in \boldsymbol{R}^{k\times h}$.

STAGE 4: Set $U = Q_1 G_1$, $V = Q_2 G_2$, $\boldsymbol{\sigma} = (\sigma_i)_{i=1,h}$.

In the cyclic reduction algorithm the inversion of the Toeplitz-like matrix A_i must be performed. This can be obtained by means of superfast algorithms in $O(m \log^2 m)$ ops or by means of the preconditioned conjugate gradient techniques. In both cases the numerical stability of performing the matrix-vector product is crucial for the reliability of the overall computation. Using orthogonal displacement representations may strongly improve numerical stability. An alternative way for Toeplitz-like matrix inversion is described in the last section of this paper and relies on Newton's iteration and on the approximate displacement rank. Even in this case, orthogonal displacement representations are fundamental for the stability of this computation. Using orthogonal displacement representations does not substantially increase the cost of the overall computation.

3.2. The case of PH/PH/1 queueing problems. For the matrix of (1.1) with blocks (1.2), we may prove that [3] at each step $j \geq 1$ of cyclic reduction it holds

$$
\begin{aligned}
A_j &= A - (I_{m_1} \otimes \boldsymbol{S}^0)\sigma_j(\boldsymbol{\alpha} \otimes I_{m_2}) - (\boldsymbol{T}^0 \otimes I_{m_2})\theta_j(I_{m_1} \otimes \boldsymbol{\beta}) \\
B_j &= -(I_{m_1} \otimes \boldsymbol{S}^0)\delta_j(I_{m_1} \otimes \boldsymbol{\beta}) \\
C_j &= -(\boldsymbol{T}^0 \otimes I_{m_2})\gamma_j(\boldsymbol{\alpha} \otimes I_{m_2}) \\
\widehat{A}_j &= A - (\boldsymbol{T}^0 \otimes I_{m_2})\theta_j(I_{m_1} \otimes \boldsymbol{\beta}),
\end{aligned}
$$

where δ_j, γ_j, σ_j, θ_j are $m_1 \times m_1$, $m_2 \times m_2$, $m_1 \times m_2$, $m_2 \times m_1$ matrices, respectively. The matrices δ_j, γ_j, σ_j, θ_j, for $j \geq 2$, are defined by the recursions

$$
\begin{aligned}
\delta_{j+1} &= \delta_j \left(\delta_1 + \begin{bmatrix} \delta_1 & \sigma_1 \end{bmatrix} D_j V_j^{-1} \begin{bmatrix} \theta_1 \\ \delta_1 \end{bmatrix}\right)\delta_j, \\
\gamma_{j+1} &= \gamma_j \left(\gamma_1 + \begin{bmatrix} \theta_1 & \gamma_1 \end{bmatrix} D_j V_j^{-1} \begin{bmatrix} \gamma_1 \\ \sigma_1 \end{bmatrix}\right)\gamma_j, \\
\sigma_{j+1} &= \sigma_j + \delta_j \left(\sigma_1 + \begin{bmatrix} \delta_1 & \sigma_1 \end{bmatrix} D_j V_j^{-1} \begin{bmatrix} \gamma_1 \\ \sigma_1 \end{bmatrix}\right)\gamma_j, \\
\theta_{j+1} &= \theta_j + \gamma_j \left(\theta_1 + \begin{bmatrix} \theta_1 & \gamma_1 \end{bmatrix} D_j V_j^{-1} \begin{bmatrix} \theta_1 \\ \delta_1 \end{bmatrix}\right)\delta_j,
\end{aligned}
$$

for $j = 1, 2, \ldots$, where

$$\begin{aligned}
\delta_1 &= (I_{m_1} \otimes \boldsymbol{\beta}) A^{-1} (I_{m_1} \otimes \boldsymbol{S}^0), \\
\gamma_1 &= (\boldsymbol{\alpha} \otimes I_{m_2}) A^{-1} (\boldsymbol{T}^0 \otimes I_{m_2}), \\
\sigma_1 &= (I_{m_1} \otimes \boldsymbol{\beta}) A^{-1} (\boldsymbol{T}^0 \otimes I_{m_2}), \\
\theta_1 &= (\boldsymbol{\alpha} \otimes I_{m_2}) A^{-1} (I_{m_1} \otimes \boldsymbol{S}^0),
\end{aligned}$$

and

$$D_j = \begin{bmatrix} \sigma_j & 0 \\ 0 & \theta_j \end{bmatrix}, \quad V_j = I - \begin{bmatrix} \theta_1 & \gamma_1 \\ \delta_1 & \sigma_1 \end{bmatrix} D_j.$$

In this way instead of processing matrices of size $m_1 m_2$, we have to deal just with matrices of size $m_1 + m_2$. The cost of one step is so reduced from $O(m_1^3 m_2^3)$ to $O(m_1^3 + m_2^3)$ ops. The algorithm has been implemented and the results of the numerical tests proved its reliability and efficiency. More details can be found in [3].

3.3. The Ramaswami reduction: Frobenius structures. One of the main problems in the solution of M/G/1 type Markov chains is the computation of the minimal nonnegative solution G of the nonlinear matrix equation $G = \sum_{i=0}^{\infty} D_i G^i$, where D_i, $i \geq 0$, are $m \times m$ nonnegative matrices such that $\sum_{i=0}^{\infty} D_i$ is stochastic. In [28], V. Ramaswami has proved that this problem is reduced to the problem of solving the infinite block tridiagonal system

$$Q = \begin{bmatrix} \widehat{A} & C & & \bigcirc \\ B & A & C & \\ & B & A & \ddots \\ \bigcirc & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} H \\ H^2 \\ H^3 \\ \vdots \end{bmatrix} = \begin{bmatrix} B \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \tag{3.2}$$

with infinite blocks defined by

$$A = \begin{bmatrix} I - D_1 & -D_2 & -D_3 & \cdots \\ & I & & \\ & & I & \\ & & & \ddots \end{bmatrix}, \quad B = \begin{bmatrix} -D_0 & 0 & \cdots \\ 0 & 0 & \cdots \\ \vdots & \vdots & \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & 0 & \cdots \\ -I & 0 & \ddots \\ & -I & \ddots \\ & & \ddots \end{bmatrix}. \tag{3.3}$$

In [12] we show that, if $D_i = 0$, for $i > N$, then the system (3.2) can be reduced to an infinite block tridiagonal system, with $N \times N$ blocks, obtained by taking finite N-dimensional sections of matrices (3.3). Moreover, under the notation of (2.14) if we apply cyclic reduction to the latter system, the sequence $W_j = -(I - \widehat{A}_j)^{-1} B$ is such that $H = W_j + O(\sigma^{2^j})$ where $0 < \sigma < 1$, and where A_0, B_0, C_0 are the N-dimensional finite sections of the matrices (3.3), and H is the N-dimensional finite section of the corresponding matrix in (3.2). The efficient computation of the blocks generated by the cyclic reduction process relies on the Frobenius structure of A_j, B_j, C_j. Indeed, it is a simple matter to show that the matrices A_j and B_j remain structured at each

step j: B_j is a matrix having all the entries equal to zero, except for the entries of the block (1,1); A_j has the structure $I + \boldsymbol{u}_j \boldsymbol{e}_1^T + \boldsymbol{e}_1 \boldsymbol{v}_j^T$, where $\boldsymbol{v}_j$, $\boldsymbol{u}_j$ are suitable block vectors, and $\boldsymbol{e}_1$ is the first block column of the identity matrix. On the other hand the matrix C_j, after $\log N$ steps is a full matrix, without any evident structure. In fact, also the matrix C_j is structured, and the structure is related to the algebra generated by a block Frobenius matrix. More precisely, if D_0 is nonsingular, define the block Frobenius matrix

$$F = \begin{bmatrix} -D_0^{-1}(D_1 - I) & -D_0^{-1}D_2 & \dots & -D_0^{-1}D_N \\ I & 0 & \dots & 0 \\ & \ddots & \ddots & \vdots \\ \bigcirc & & I & 0 \end{bmatrix}.$$

Then the matrix polynomial $\phi(z) = B + zA + z^2C$ can be written as

$$\phi(z) = \begin{bmatrix} D_0 & & & \bigcirc \\ & -zI & & \\ & & \ddots & \\ \bigcirc & & & -zI \end{bmatrix} (I - zF)$$

that is, $\phi(z)$ is a block diagonal matrix times a matrix polynomial with block coefficients defined by powers of F. A similar structure is kept at each step j; more precisely, we prove that the matrix polynomial of (2.15) $\phi_j(z) = C_j + zA_j + z^2B_j$ can be expressed as

$$\phi_j(z) = (I - \boldsymbol{w}_j(z) H_j \boldsymbol{e}_1^T)(zI - F^{2^j}) \tag{3.4}$$

where $H_j = (\boldsymbol{e}_1^T F^{2^j-1} \boldsymbol{e}_1)^{-1}$, $\boldsymbol{w}_j(z) = F^{2^j-1}\boldsymbol{e}_1 + z\boldsymbol{e}_1 D_0$. That is, up to a small rank correction, $\phi_j(z)$ is a matrix polynomial with block coefficients generated by powers of F. From this property it follows that $B_j = \phi_j(0)$ has displacement rank at most 2. By using a continuity argument, we prove that the same bound holds also in case where D_0 is singular.

Thus, the computation of A_j, B_j, C_j costs $O(m^3 N + m^2 N \log N)$ ops.

3.4. The case of block banded block Toeplitz matrices. Consider the block banded block Toeplitz matrix Q with banded Toeplitz blocks, described in section 1.3. By partitioning Q into $(mM) \times (mM)$ blocks we obtain a block tridiagonal matrix with blocks A, B, C. It is immediate to verify that these blocks have displacement rank at most $2M$. In particular, for the matrix polynomial $\phi(z)$ and for the matrix function $\psi(z) = \phi(z)^{-1}$ it holds $\mathrm{Drk}(\phi(z)) = \mathrm{Drk}(\psi(z)) \leq 2M$, for any z.

The matrix polynomial $\phi(z)$ has no algebraic properties that could lead to a conservation of the displacement rank. Even in the case $M = 1$, where $\phi(z)$ is a tridiagonal Toeplitz matrix, the matrix polynomials $\phi_j(z)$ do not maintain a constant displacement rank.

A possible way for overcoming this drawback is to look for different displacement operators associated with different matrix algebras like trigonometric algebras. This is a research direction to be better investigated.

A different approach is based on the concept of approximate displacement rank introduced in section 2.3. Indeed, under the assumptions that guarantee the convergence to zero of the blocks C_j, B_j (compare with Theorems 2.8, 2.9), it is easy to prove

that for a given ϵ, $\mathrm{Drk}_\epsilon(A_j)$ is bounded by a constant independent of j. Similarly it holds for $\mathrm{Rdrk}_\epsilon(B_j)$, and $\mathrm{Rdrk}_\epsilon(C_j)$.

The following modification of cyclic reduction that generates approximations $\widetilde{A}_j$, $\widetilde{B}_j$, $\widetilde{C}_j$, $\widetilde{\widehat{A}}_j$, of the blocks A_j, B_j, C_j, $\widehat{A}_j$, expressed in terms of their relative ϵ-displacement generator is based on the above property.

For simplicity we just give an outline of the algorithm in its general j-th step.

ALGORITHM 3.4.

INPUT: A relative ϵ-displacement generator of $\widetilde{A}_j$, $\widetilde{B}_j$, $\widetilde{C}_j$, $\widetilde{\widehat{A}}_j$.

OUTPUT: A relative ϵ-displacement generator of $\widetilde{A}_{j+1}$, $\widetilde{B}_{j+1}$, $\widetilde{C}_{j+1}$, $\widetilde{\widehat{A}}_{j+1}$.

COMPUTATION:

1. Compute the first column $\widetilde{\boldsymbol{b}} = -\widetilde{B}\widetilde{A}^{-1}\widetilde{B}\boldsymbol{e}_1$.
2. By using the relative ϵ-displacement generator of $\widetilde{A}_j$, compute a relative δ-displacement generator of $\widetilde{A}_j^{-1}$, for $\delta = \epsilon\mathrm{Cond}(\widetilde{A}_j)$.
3. From the latter, and from the relative ϵ-displacement generator of $\widetilde{B}_j$ compute a relative θ-displacement generator of
$$H = -\widetilde{B}_j\widetilde{A}_j^{-1}\widetilde{B}_j,$$
for a suitable θ, according to formulae (2.8).
Let h be the rank of such generator and S_1 and S_2 be the $mM \times h$ matrices whose columns define a generator, i.e., such that
$$\Delta(-\widetilde{B}_j\widetilde{A}_j^{-1}\widetilde{B}_j) = S_1S_2^T + E, \quad ||E|| \le \theta||\widetilde{B}_j\widetilde{A}_j^{-1}\widetilde{B}_j||.$$
4. Compute the QR factorizations $S_1 = Q_1R_1$, $S_2 = Q_2R_2$ and the $h \times h$ matrix $W = R_1R_2^T$. Compute the singular value decomposition $W = U\Sigma V^T$ of W and denote $\sigma_1, \ldots, \sigma_h$ the singular values of W.
Let r be such that $\sigma_r > \epsilon||\widetilde{B}_j\widetilde{A}_j^{-1}\widetilde{B}_j|| \ge \sigma_{r+1}$ and denote with $\widehat{U}$, $\widehat{V}$, the matrices made by the first r columns of U and V, respectively. Similarly denote with $\widehat{\Sigma}$ the leading $r \times r$ principal submatrix of Σ.
5. Define the matrix $\widetilde{B}_{j+1}$ as the one represented by the displacement generator $(\widetilde{\boldsymbol{b}}, Q_1\widehat{U}, \widehat{\boldsymbol{\sigma}}, Q_2\widehat{V})$, $\widehat{\boldsymbol{\sigma}} = (\sigma_1, \ldots, \sigma_r)$, i.e., $\widetilde{B}_{j+1}$ is the matrix such that $\Delta(\widetilde{B}_{j+1}) = (Q_1\widehat{U})\widehat{\Sigma}(Q_2\widehat{V})^T$. This matrix provides a relative ϵ-displacement representation of H.
6. Perform similar computations for the matrices $-\widetilde{C}_j\widetilde{A}_j^{-1}\widetilde{C}_j$, $\widetilde{A}_j - \widetilde{B}_j\widetilde{A}_j^{-1}\widetilde{C}_j - \widetilde{C}_j\widetilde{A}_j^{-1}\widetilde{B}_j$ and for $\widetilde{\widehat{A}}_j - \widetilde{B}_j\widetilde{A}_j^{-1}\widetilde{C}_j$.

The cost of the algorithm is $O(h^2mM\log^2(mM))$ ops per step and strongly depends on the largest values of h encountered in the computation. The algorithm has been implemented and tested with certain PSF's like the one defined by $M = 4$, $\alpha_{0,0} = \gamma/\sigma$, $\alpha_{i,j} = 1/\sigma$ otherwise, with the normalizing factor $\sigma = (M+1)^2 - 1 + \gamma$. For $\gamma > 2$ the algorithm has been capable to reconstruct the image with a good accuracy even with rather large values of ϵ, say in the range $[10^{-6}, 10^{-2}]$.

Issues like the numerical stability and the control of the level ϵ where the singular values are truncated are still to be investigated.

The following tables report the growth of the (relative) displacement rank of the blocks A_j, B_j, C_j and $\widehat{A}_j$ generated by Algorithm 3.4 when applied to the block 5-diagonal block Toeplitz matrix having Toeplitz blocks with 5 diagonals, where all the nonzero off-diagonal entries are 1 and the diagonal entry is α.

	$\epsilon = 10^{-2}$		$\epsilon = 10^{-4}$		$\epsilon = 10^{-6}$	
m	Drk_ϵ	Rdrk_ϵ	Drk_ϵ	Rdrk_ϵ	Drk_ϵ	Rdrk_ϵ
64	12	12	20	22	28	34
128	12	12	20	22	28	34
256	12	12	20	22	28	34

TABLE 3.1
$\alpha = 8$

m	Drk_ϵ	Rdrk_ϵ
64	52	22
128	88	42
256	150	62

TABLE 3.2
$\alpha = 2, \epsilon = 10^{-2}$

In tables 3.1, 3.2 we report the size m of the blocks of the reblocked tridiagonal matrix and the largest values of Drk_ϵ and Rdrk_ϵ of the blocks generated by CR, for different values of ϵ. We observe that for $\alpha = 2$ the hypotheses of theorem 2.8 are not satisfied and even though the Rdrk_ϵ depends on N, its growth is rather moderate.

3.5. Inverting Toeplitz-like matrices with Newton's iteration. Newton's iteration is a valid tool for solving matrix problems where the solution to be computed can be formally viewed as the zero of a known function. For instance, the formal application of this iteration to the matrix equation $X^{-1} - A = 0$ yields

$$X_{i+1} = 2X_i - X_i A X_i, \quad i \geq 0 \tag{3.5}$$

that for a suitable X_0 provides a sequence converging to A^{-1} if $\det A \neq 0$. This method has been analyzed in [13] and, concerning Toeplitz matrices, has been investigated in [27] where the iteration (3.5) has been modified by replacing each X_i with the closest matrix having displacement rank 2. Conditions on X_0 have been given in order that the method, modified in this way, still converges.

Observe that for a nonsingular $n \times n$ Toeplitz matrix A, its inverse A^{-1} has displacement rank at most 2, whereas the matrices X_i generated by Newton's iteration have displacement rank that generally grows and after few steps coincides with n. Despite the displacement rank of X_i is large, the Euclidean distance of X_i from the set of matrices with displacement rank at most 2 tends quadratically to zero. This is a crucial fact that suggests us to apply the concept of approximate displacement rank.

Here we recall the different approach presented in [9], for inverting a Toeplitz-like matrix A having displacement rank k, based on the idea of *adaptivity*, where each approximation X_i is replaced by the closest matrix having displacement rank k_i, and k_i is chosen so that the Euclidean distance of X_{i+1} from the matrices with displacement rank k is "sufficiently" small in order to guarantee the convergence.

In this way, the displacement rank is not bounded a priori but it may vary according to the difficulty of the problem and to the needs of the computation. This approach is adaptive in the sense that the cost of the algorithm is not uniform with respect to the set of input data, where the algorithm adjusts the computation to the specific difficulty of the problem. In practice, for well conditioned matrices the

growth of the ϵ-displacement rank is negligible and the overall cost is $O(n \log n)$ ops. For (polynomially) ill-conditioned matrices the approximate displacement rank seems to grow as $O(\log n)$ so that the overall cost is $O(n \log^2 n)$ ops.

We refer the reader to [9] for all the issues concerning convergence, preconditioning, regularization properties and generalizations of Newton's iterations. Here we just want to describe how approximate displacement rank can be combined with Newton's iteration in order to increase the performances of this method if applied to Toeplitz-like matrices.

Let us assume that $\mathrm{Drk}(A) = k$ and consider the following modification of (3.5)

$$\begin{aligned} Y_i &= 2X_i - X_i A X_i \\ X_{i+1} &= \mathrm{Trunc}_{\epsilon_i}(Y_i), \quad i \geq 0, \end{aligned} \tag{3.6}$$

where $\mathrm{Trunc}_{\epsilon_i}(Y_i)$ is the relative ϵ_i-displacement generator of Y_i given according to (2.6), and ϵ_i is a parameter that is dynamically computed during the iteration. The value of ϵ_i should be not too small to allow the displacement rank of X_i to reach large values, and, at the same time, should be not too large to spoil the convergence of the sequence X_i.

The choice of ϵ_i can be made according to different strategies. The simplest one is to choose a constant value of ϵ_i, say, the square root of the machine precision of the computation. A dynamic choice of ϵ that guarantees the convergence of the modified Newton method can be easily designed. We report below a convergence theorem, here adjusted to the operator $\Delta(A) = ZA - AZ$ and originally proved in [9] for a slightly different operator.

THEOREM 3.5. *Let X_0 be such that $||X_0|| \leq 1-\theta$, set $\epsilon_i = \theta/(4n(2h_i+k)\sigma_1^{(i)}||A||)$, $h_i = \mathrm{Drk}(X_i)$, $\sigma_1^{(i)} = \rho(\Delta(X_i))$, $k = \mathrm{Drk}(A)$. Then for the residuals $R_i = I - AX_i$ of the sequence $\{X_i\}_i$ generated by the modified Newton iteration (3.6) it holds $||R_i|| \leq (1 - \theta/2)^{2^i}$.*

For the computation of a single step of modified Newton's iteration observe that

$$\Delta(Y_i) = 2\Delta(X_i) - \Delta(X_i)AX_i - X_i\Delta(A)X_i - X_iA\Delta(X_i) \tag{3.7}$$

so that $\mathrm{Drk}(Y_i) \leq k + 2\mathrm{Drk}(X_i)$ where $k = \mathrm{Drk}(A)$.

Moreover, replacing for notational simplicity X_i with X and Y_i with Y, if $\Delta(X) = U_X D_X V_X^T$, $\Delta(A) = U_A D_A V_A^T$, where $U_X, V_X, U_A, V_A \in \boldsymbol{R}^{n \times h}$ are orthogonal, we obtain that

$$\begin{aligned} &\Delta(Y) = U_Y D_Y V_Y^T \\ &U_Y = [U_X, XU_A, XAU_X], \quad V_Y = [V_X, X^T V_A, X^T A^T V_X] \\ &D_Y = \begin{bmatrix} 2D_X & 0 & -D_X \\ 0 & -D_A & 0 \\ -D_X & 0 & 0 \end{bmatrix} \end{aligned} \tag{3.8}$$

Thus, an orthogonal generator of $X_{i+1} = \mathrm{Trunc}_\epsilon(Y)$ can be computed according to the following

ALGORITHM 3.6. *Computing the approximate orthogonal generator of $Y = 2X - XAX$.*

1. Compute the first column of X_{i+1}, i.e., $2X_i\boldsymbol{e}_1 - X_iAX_i\boldsymbol{e}_1$;
2. compute the matrices U_Y, V_Y, D_Y of (3.8);
3. compute the QR factorizations $U_Y = Q_1R_1$, $V_Y = Q_2R_2$;

4. compute the SVD $R_1 D_Y R_2^T = U\Sigma V$, and r such that for the singular values of $R_1 D_Y R_2^T$ it holds $\sigma_{r+1} \leq \epsilon\sigma_1 < \sigma_r$;
5. set $U_X = Q_1\widehat{U}$, $V_X = Q_2\widehat{V}$, $\boldsymbol{d}_X = (\sigma_1, \ldots, \sigma_r)$, where $\widehat{U}$ and $\widehat{V}$ are made up by the first r columns of U and V, respectively.

The cost of the above computation is about $O(hk)$ FFTs. Improvements of this computation can be obtained by suitably using further structures, like the persymmetry (as in the case of Toeplitz matrices) or the symmetry of the matrices involved, that allow the reduction of the number of FFTs.

Below we resume our modification of Newton's iteration for the inversion of a Toeplitz-like matrix assigned in terms of its orthogonal displacement generator.

ALGORITHM 3.7. *Newton's iteration with ϵ-displacement rank.*

INPUT Integers n, k, the maximum number $MAXIT$ of iterations, a residual error bound ϵ, the orthogonal generator $(\boldsymbol{a}, U_A, \boldsymbol{\sigma}_A, V_A)$ of the nonsingular matrix $A \in \boldsymbol{R}^{n\times n}$ having displacement rank k; the orthogonal generator $(\boldsymbol{x}, U_X, \boldsymbol{\sigma}_X, V_X)$ of a matrix X_0 such that AX_0 is positive definite and $\rho(AX_0) < 1$; a strategy for the choice of the truncation value ϵ_i.

OUTPUT An approximation X of A^{-1} such that $||I - AX|| \leq \epsilon$ given in terms of its orthogonal generator $(\boldsymbol{x}, U_X, \boldsymbol{\sigma}_X, V_X)$.

COMPUTATION

STAGE 0 Set $i = 0$, $X = X_0$;

STAGE 1 Compute the truncation value ϵ_i;

STAGE 2 Compute the approximate orthogonal generator $(\boldsymbol{y}, U_Y, \boldsymbol{\sigma}_Y, V_Y)$ of $Y = 2X - XAX$, by means of algorithm 3.6 with $\epsilon = \epsilon_i$ and set $(\boldsymbol{x}, U_X, \boldsymbol{\sigma}_X, V_X) = (\boldsymbol{y}, U_Y, \boldsymbol{\sigma}_Y, V_Y)$

STAGE 3 Compute $||R||$ for the residual $R = I - AX$; if $||R|| > \epsilon$ and $i \leq MAXIT$ set $i = i + 1$ and continue from Stage 1.

STAGE 4 If $i > MAXIT$ then output FAILURE, otherwise output the orthogonal generator of X.

Concerning the choice of ϵ_i, Theorem 3.5 provides the strategy that guarantees convergence. Different strategies based on heuristics can be applied. The starting matrix X_0 can be chosen according to the features of A: for a general A, we may set $X_0 = A^T/||AA^T||$; for A symmetric and positive definite we may set $X_0 = I/||A||$. A more convenient choice can be done in terms of the least square circulant matrix that approximates A.

3.5.1. Numerical experiments. We have applied the modified Newton iteration for inverting some test Toeplitz matrices. In Tables 3.3, 3.4, 3.5, 3.6 we report the number of steps $Nsteps$, the maximum displacement rank $Mdrk$, the sum of the displacement ranks $Sdrk$ of each step, and the residual error err for inverting the matrix A_n, for each value of n, where $A_n = (a_{|i-j|})_{i,j=1,n}$ is defined in the following way:

MATRIX 1 **of Table 3.3:** $a_{|i-j|} = 1/(|i-j|+1)$;
MATRIX 2 **of Table 3.4:** $a_0 = 4, a_1 = 1, a_i = 0$ otherwise;
MATRIX 3 **of Table 3.5:** $a_0 = 2.1, a_1 = 1, a_i = 0$ otherwise;
MATRIX 4 **of Table 3.6:** $a_0 = 2, a_1 = 1, a_i = 0$ otherwise.

The quantity $Sdrk \cdot n \log n$ expresses the overall arithmetic cost of the algorithm.

The matrices 1,2,3 are asymptotically well conditioned while the matrix 4 has condition number that grows as $O(n^2)$.

n	$Nstep$	$Mdrk$	$Sdrk$	err
32	8	3	16	2.4e-14
64	8	3	16	5.1e-14
128	8	3	16	6.6e-14
256	8	3	16	6.8e-14
512	8	3	16	6.0e-14
1024	8	3	16	5.8e-14
2048	8	3	16	5.3e-14
4096	8	3	16	4.7e-14

TABLE 3.3

n	$Nstep$	$Mdrk$	$Sdrk$	err
32	8	3	15	2.0e-16
64	8	3	15	4.0e-16
128	8	3	15	2.0e-16
256	8	3	15	2.0e-16
512	8	3	15	2.0e-16
1024	8	3	15	2.1e-16
2048	8	3	15	2.0e-16
4096	8	3	15	2.2e-16

TABLE 3.4

n	$Nstep$	$Mdrk$	$Sdrk$	err
32	11	4	38	3.3e-14
64	12	5	41	1.7e-15
128	11	8	52	2.0e-15
256	11	7	47	2.1e-15
512	10	6	41	3.4e-13
1024	10	5	38	2.5e-13
2048	10	5	35	1.9e-13
4096	10	4	29	2.2e-13

TABLE 3.5

n	$Nstep$	$Mdrk$	$Sdrk$	err
32	12	5	47	3.8e-15
64	13	7	57	2.5e-15
128	14	8	68	8.9e-15
256	15	8	76	4.8e-14
512	16	10	93	6.5e-14
1024	18	13	121	1.1e-13
2048	22	19	187	3.2e-13
4096	23	22	224	1.3e-13

TABLE 3.6

For well conditioned matrices the number of iterations as well as the maximum displacement rank is independent of n. For the matrix 4, the number of iterations as well as the sum of the displacement ranks seems to grow as $\log_2 n$.

Acknowledgments. This work has been supported by the MURST project "Advanced numerical methods for scientific computing" and by the Coordinated CNR project "Toeplitz-like matrices: structures algorithms and applications".

[1] P. Amodio and F. Mazzia. Backward error analysis of cyclic reduction for the solution of tridiagonal systems. *Math. Comp.*, 62:601–617, 1994.

[2] D. A. Bini. Using FFT-based techniques in polynomial and matrix computations: recent advances and applications. In *Proceedings of FAA98, Kuwait*, 1998. To appear.

[3] D. A. Bini, S. Chakravarthy, and B. Meini. A new algorithm for the design of finite capacity service units. In B. Plateau, W. J. Stewart, and M. Silva, editors, *Numerical Solution of Markov Chains, 3rd International Workshop, Saragoza (Spain), Sept. 1999*, pages 247–260. Prensas Unversitarias de Zaragoza, 1999.

[4] D. A. Bini, L. Gemignani, and B. Meini. Factorization of Analytic Functions by means of Koenig's Theorem and Toeplitz Computations. Submitted for publication, 1998.

[5] D. A. Bini and B. Meini. On cyclic reduction applied to a class of Toeplitz-like matrices arising in queueing problems. In W. J. Stewart, editor, *Computations with Markov Chains*, pages 21–38. Kluwer Academic Publisher, 1995.

[6] D. A. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM J. Matrix Anal. Appl.*, 17:906–926, 1996.

[7] D. A. Bini and B. Meini. Improved cyclic reduction for solving queueing problems. *Numerical Algorithms*, 15:57–74, 1997.

[8] D. A. Bini and B. Meini. Using displacement structure for solving non-skip-free M/G/1 type Markov chains. In A. Alfa and S. Chakravarthy, editors, *Advances in Matrix Analytic Methods for Stochastic Models - Proceedings of the 2nd international conference on matrix analytic methods*, pages 17–37. Notable Publications Inc, NJ, 1998.

[9] D. A. Bini and B. Meini. Approximate displacement rank and applications. In V. Olshevsky, editor, *Proceedings AMS Conference "Structured Matrices in Operator Theory, Numerical Analysis, Control, Signal and Image Processing", Boulder*. AMS, 1999. to apppear.

[10] D. A. Bini and B. Meini. Effective methods for solving banded Toeplitz systems. *SIAM J. Matrix Anal. Appl.*, 20:700–719, 1999.

[11] D. A. Bini and B. Meini. Solving block banded block Toeplitz systems with banded Toeplitz blocks. In F. T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures, and Implementations IX*, volume 3807, pages 300–311. Proceedings of SPIE, 1999.

[12] D. A. Bini, B. Meini, and V. Ramaswami. Analyzing M/G/1 paradigms through QBDs: the role of the block structure in computing the matrix G. submitted for publication, 1999.

[13] D. A. Bini and V. Pan. *Matrix and Polynomial Computations, Vol. 1: Fundamental Algorithms*. Birkhäuser, Boston, 1994.

[14] R. R. Bitmead and B. D. O. Anderson. Asymptotically fast solution of Toeplitz and related systems of linear equations. *Linear Algebra Appl.*, 34:103–116, 1980.

[15] H. R. Gail, S. L. Hantler, and B. A. Taylor. Solutions of the basic matrix equation for M/G/1 and G/M/1 type Markov chains. *Commun. Statist. Stochastic Models*, 10:1–43, 1994.

[16] H. R. Gail, S. L. Hantler, and B. A. Taylor. Non-skip-free M/G/1 and G/M/1 type Markov chains. *Adv. Appl. Prob.*, 29:733–758, 1997.

[17] G. Heinig and K. Rost. *Algebraic Methods for Toeplitz-like Matrices and Operators*. Akademie-Verlag, Berlin, and Birkhäuser, Boston, 1984.

[18] Don Heller. Some aspects of the cyclic reduction algorithm for block tridiagonal linear systems. *SIAM J. Num. Anal.*, 13:484–496, 1976.

[19] A. S. Householder. *The Numerical Treatment of a Single Nonlinear Equation*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, NY, 1970.

[20] T. Kailath. Displacement structure and array algorithms. In T. Kailath and A. H. Sayed, editors, *Fast reliable algorithms for matrices with structure*, chapter 1, pages 1–56. SIAM, Philadelphia, 1999.

[21] T. Kailath, S. Kung, and M. Morf. Displacement rank of a matrix. *Bulletin of the American Mathematical Society*, 1:769–773, 1979.

[22] T. Kailath, S. Kung, and M. Morf. Displacement ranks of matrices and linear equations. *J. Math. Anal. Appl.*, 68:395–407, 1979.

[23] P. Lancaster and M. Tismenetski. *Matrix Polynomials*. Academic Press Inc., New York, 1982.

[24] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for Quasi-Birth-Death processes. *J. Appl. Probability*, 30:650–674, 1993.

[25] B. Meini. Efficient computation of the extreme solutions of $X + A^*X^{-1}A = Q$, $X - A^*X^{-1}A = Q$. Submitted for publication, 1999.

[26] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, Baltimore, MD, 1981.

[27] V. Y. Pan, S. Branham, R. E. Rosholt, and A.-L. Zheng. Newton's iteration for structured

matrices. In T. Kailath and A. H. Sayed, editors, *Fast reliable algorithms for matrices with structure*, chapter 7, pages 189–210. SIAM, Philadelphia, 1999.

[28] V. Ramaswami. The generality of Quasi Birth-and-Death processes. In A. Alfa and S. Chakravarthy, editors, *Advances in Matrix Analytic Methods for Stochastic Models - Proceedings of the 2nd international conference on matrix analytic methods*, pages 93–113. Notable Publications Inc, NJ, 1998.

[29] P. Tilli. Locally Toeplitz sequences: spectral properties and applications. *Linear Algebra Appl.*, 278:91–120, 1998.

[30] P. Tilli. Asymptotic spectral distribution of Toeplitz-related matrices. In T. Kailath and A. H. Sayed, editors, *Fast reliable algorithms for matrices with structure*, chapter 6, pages 153–187. SIAM, Philadelphia, 1999.

[31] P. Yalamov and V. Pavlov. Stability of the block cyclic reduction. *Linear Algebra Appl.*, 249:341–358, 1996.